Research Article | Vol. 11, No. 4 / 1 April 2020 / Biomedical Optics Express 1876
Biomedical Optics EXPRESS

Detail-preserving pulse wave extraction from facial videos using consumer-level camera

DINGLIANG WANG,[1] XUEZHI YANG,[2,3,*] XUENAN LIU,[1] JIN JING,[1] AND SHUAI FANG[1]

[1]School of Computer and Information, Hefei University of Technology, Hefei, 230009, China
[2]School of Software, Hefei University of Technology, Hefei, 230009, China
[3]Anhui Key Laboratory of Industry Safety and Emergency Technology, Hefei, 230009, China
[*]xzyang@hfut.edu.cn

Abstract: With the popularity of smart phones, non-contact video-based vital sign monitoring using a camera has gained increased attention over recent years. Especially, imaging photoplethysmography (IPPG), a technique for extracting pulse waves from videos, conduces to monitor physiological information on a daily basis, including heart rate, respiration rate, blood oxygen saturation, and so on. The main challenge for accurate pulse wave extraction from facial videos is that the facial color intensity change due to cardiovascular activities is subtle and is often badly disturbed by noise, such as illumination variation, facial expression changes, and head movements. Even a tiny interference could bring a big obstacle for pulse wave extraction and reduce the accuracy of the calculated vital signs. In recent years, many novel approaches have been proposed to eliminate noise such as filter banks, adaptive filters, Distance-PPG, and machine learning, but these methods mainly focus on heart rate detection and neglect the retention of useful details of pulse wave. For example, the pulse wave extracted by the filter bank method has no dicrotic wave and approaching sine wave, but dicrotic waves are essential for calculating vital signs like blood viscosity and blood pressure. Therefore, a new framework is proposed to achieve accurate pulse wave extraction that contains mainly two steps: 1) preprocessing procedure to remove baseline offset and high frequency random noise; and 2) a self-adaptive singular spectrum analysis algorithm to obtain cyclical components and remove aperiodic irregular noise. Experimental results show that the proposed method can extract detail-preserved pulse waves from facial videos under realistic situations and outperforms state-of-the-art methods in terms of detail-preserving and real time heart rate estimation. Furthermore, the pulse wave extracted by our approach enabled the non-contact estimation of atrial fibrillation, heart rate variability, blood pressure, as well as other physiological indices that require standard pulse wave.

© 2020 Optical Society of America under the terms of the OSA Open Access Publishing Agreement

1. Introduction

Long-term and regular monitoring of vital signs such as heart rate (HR), heart rate variability (HRV), blood oxygen saturation (SpO2), blood viscosity and blood pressure (BP) are important for the early warning of cardiovascular diseases. Currently, the gold standard techniques to measure the vital signs are based on contact sensors or specific devices such as electrocardiogram (ECG) probes, blood pressure cuffs and pulse oximeters. However, these sensors or devices are inconvenient for daily use and may cause skin irritation in case of treatment for newborns.

In the past decade, researchers have focused on non-contact detection methods and found that the change of blood volume in the capillaries underneath the skin leads to small change in the skin color which can be captured by consume-level cameras. This discovery directly motivated the development of imaging photoplethysmography (IPPG), a technique for extracting pulse waves from videos [1–3]. In 2011, Poh et al. first realized facial heart rate detection using a webcam [11], since then, non-contact heart rate detection based on IPPG technology has become a research hotspot.

#380646 https://doi.org/10.1364/BOE.380646
Journal © 2020 Received 15 Oct 2019; revised 29 Jan 2020; accepted 4 Feb 2020; published 11 Mar 2020

Corrected: 03 Apr 2020

Biomedical Optics EXPRESS

The main challenge for accurate pulse wave extraction from facial videos is that the color change of the face due to cardiovascular activities is subtle and many factors could contaminate the pulse wave and make the extraction task difficult, for example, both environmental illumination variations and subjects' motions could affect the gray value of the face region significantly, such as the flicker of light, the inner noise of the digital camera, the facial expression changes and the unconsciously shaking of head. Recent years, many new methods have been proposed to detect pulse wave which are mainly focused on HR estimation and overlooked the preservation of pulse details [9–15]. Yet, effective methods for facial videos under realistic situations are scarce. Li et al. proposed an anti-interference method called normalized Least Mean Square (NLMS) adaptive filter which could rectify the illuminance variation [16], but a smooth rectifier in background is critical for establishing the desired signal as the input of the NLMS adaptive filer, which is hard to realize practically. Afterwards, some novel methods were proposed, in 2014, Yu et al. proposed the filter bank method and achieved satisfying accuracy of HR estimation [17], this method divides the bandwidth from 0.8 Hz to 4 Hz into 16 sub-bands (0.02 Hz for each sub-band), then use the sub-bands to bandpass the source signal and get 16 filtered components, finally, the component with the highest energy is selected as pulse wave and HR can be estimated according to the central frequency of the corresponding sub-band. Due to the usage of 0.02Hz narrowband, the pulse wave extracted by filter bank approaching sine wave and cannot be used for the measurement of many vital signs such as blood viscosity. In 2015, Kumar et al. proposed the Distance-PPG method which combines skin-color change signals from different tracked regions of the face using a weighted average, where the weights depend on the blood perfusion and incident light intensity in the region [18]. Distance-PPG method has excellent anti-noise performance, but the pulse wave extracted by Distance-PPG could not see obvious dicrotic wave. Another shortage is that Distance-PPG algorithm is time-consuming which makes it unsuitable for real-time vital signs calculation on smart phones. To the best of our knowledge, no method has been demonstrated to be able to extract detail-reserved standard pulse wave from facial videos under realistic conditions.

In this paper, a new method is proposed to realize detail-preserving pulse wave extraction from facial videos under realistic situations. Firstly, the subject's facial region is tracked to remove rigid motion disturbance, and the original pulse wave is obtained by calculating the mean value of green channel pixels in the facial region. Afterwards, baseline cancelation, spike smoothing and five-point cubic smoothing algorithms are performed to remove baseline offset and high-frequency random noise. The self-adaptive singular spectrum analysis algorithm is then adopted to remove irregular noise caused by non-rigid motions and illumination changes while keep useful details reserved. Experimental results show that the proposed method outperforms state-of-the-art method in real time HR (RTHR) estimation and detail-preserving. Furthermore, our method has already been transplanted to smart phone and realized very accurate HR calculation. Our fitness diagnosis APP (android version) will be released in the near future.

2. Method

Pulse wave is one of the most important human body signals which contains abundant physiological information and can be used for diagnosing cardiovascular diseases or mental sickness [4–7]. As shown in Fig. 1, the pulse wave features such as ascending limb, descending limb and dicrotic wave are important for assessing cardiovascular condition [8], for example, the gradient of ascending limb is closely related to systolic blood pressure, and the sharpness of dicrotic wave is associated with arterial stiffness. Therefore, accurate pulse wave extraction from facial videos with useful details reserved is crucial for non-contact vital sign monitoring and is also the focus of this paper.

The overall framework of the proposed method is shown in Fig. 2 with an example of low signal to noise ratio (SNR). This section will explain our method in detail to show how it works.

Biomedical Optics EXPRESS

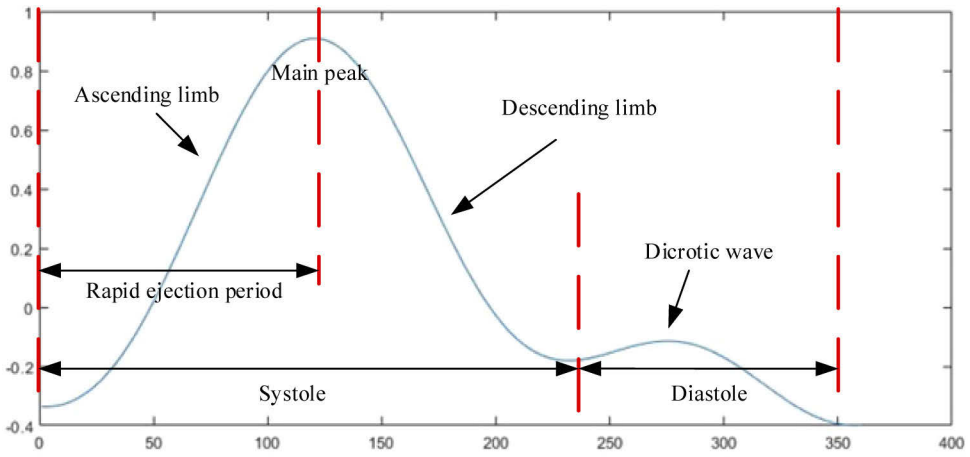

Fig. 1. Typical pulse wave of a single period.

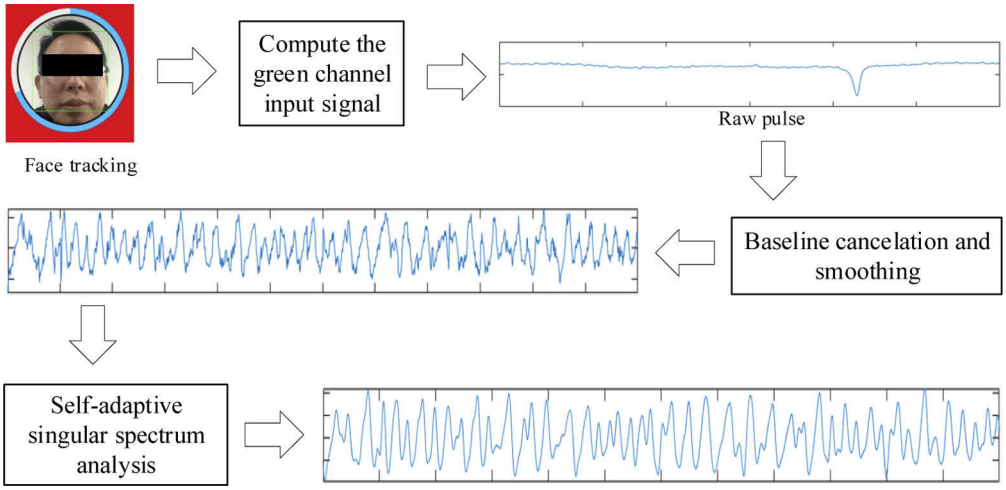

Fig. 2. Overall framework of the proposed method. The raw pulse example shown above exists a sharp drop which brings big obstacle for accurate pulse wave extraction.

2.1. Raw pulse acquisition

Face detecting and tracking are essential for raw pulse acquisition as head movements and background noise could affect pulse wave significantly. Hence the Viola-Jones face detector of OpenCV is adopted to detect the face rectangle for each video frame [20] and only the pixels in the face rectangle participating the construction of raw pulse.

The selection of color channel is crucial too. There are mainly three color models applied in raw pulse acquisition: Red-Green-Blue (RGB), Hue-Saturation-Intensity (HSI), and YCbCr where Y stands for luminance component and Cb and Cr refer to blue-difference and red-difference chroma components respectively. For HSI model, only the H channel can be used for pulse extraction and it is motion sensitive. According to Sahindrakar et al., YCbCr produces better results in detecting HR than HSI with limited movements [21]. Among these three models, the most robust model is still RGB and the green channel of RGB gives the strongest signal-to-noise

Biomedical Optics EXPRESS

ratio (SNR) [22–25]. Therefore, the raw pulse is constructed by calculating the mean value of the green channel values of pixels in the detected face rectangle.

2.2. Preprocessing

Even if the face region is tracked precisely, it may still be challenging to extract clean pulse wave during motion. As shown in Fig. 3(a), the average intensity within the facial region could experience a significant sudden drop or rise and the cause of such a shift is often due to facial expression changes or a motion-induced camera focus change [26]. To address this problem, several useful algorithms are applied including baseline cancelation, spike smoothing and five-point cubic smoothing.

1) Baseline cancelation: This algorithm is mainly used for eliminating the trend and sharp drop or rise in raw pulse. Firstly, the raw pulse is cut into several M-point non-overlapping segments and calculate the mean value of each segment. Then, detrend the signal by making the mean value of each segment equals to the mean value of overall raw pulse. According to our experiments, it is suggested to use signal sampling rate as the value of segment length M. The signal after baseline cancelation is shown as Fig. 3(b).

2) Spike smoothing: Although the signal after baseline cancelation is an approximation to the ideal pulse signal, one spike noise, which comes from the demean of sharp drop, exists. Hence the spike smoothing algorithm is proposed to deal with it. First, the raw pulse is cut into several M-point non-overlapping segments and the standard deviation of each segment is calculated. If the standard deviation of a segment exceeded two times of overall standard deviation, the corresponding segment is considered highly noisy and a compression method is employed to smooth the segment. Equation (1) is the core of the compression method:

$$CR = 0.6 + 0.4 * \frac{sd_i - 2 * sd}{sd_i} \tag{1}$$

where CR is the compress rate used to smooth the noisy segment, sd_i is the standard deviation of the noisy segment, and sd is the overall standard deviation of the pulse signal. It is obvious that the calculated CR increases with the increase of sd_i. All pulse wave values in the noisy segment are multiplied by (1- CR) to achieve spike smoothing, and the actual effect of the proposed spike smoothing algorithm is shown in Fig. 3(c).

3) Five-point cubic smoothing: This algorithm is used to remove high frequency random noise while keep the original curve characteristics unchanged. Firstly, a cubic polynomial is used to fit experimental data:

$$Y(t) = a_0 + a_1 t + a_2 t^2 + a_3 t^3 \tag{2}$$

Then compute the undetermined coefficients in Eq. (2) by defining a least square formula shown below:

$$\sum_{i=-2}^{2} R_i^2 = \sum_{i=-2}^{2} \left[\sum_{j=0}^{3} a_j t_i^j - Y_i \right]^2 = \phi(a_0, a_1, a_2, a_3) \tag{3}$$

To make $\phi(a_0, a_1, a_2, a_3)$ minimum, we compute partial derivative for a_k (k = 0,1,2,3) and let the partial derivative be zero, the following equation will be obtained:

$$\sum_{i=-2}^{2} Y_i t_i^k = \sum_{i=-2}^{2} t_i^k \sum_{j=0}^{3} a_j t_i^j \tag{4}$$

The undetermined coefficients a_k (k = 0, 1, 2, 3) can be calculated from Eq. (4), and the five-point cubic smoothing formulas below are obtained by substituting a_k (k = 0, 1, 2, 3) into Eq. (2).

Biomedical Optics EXPRESS

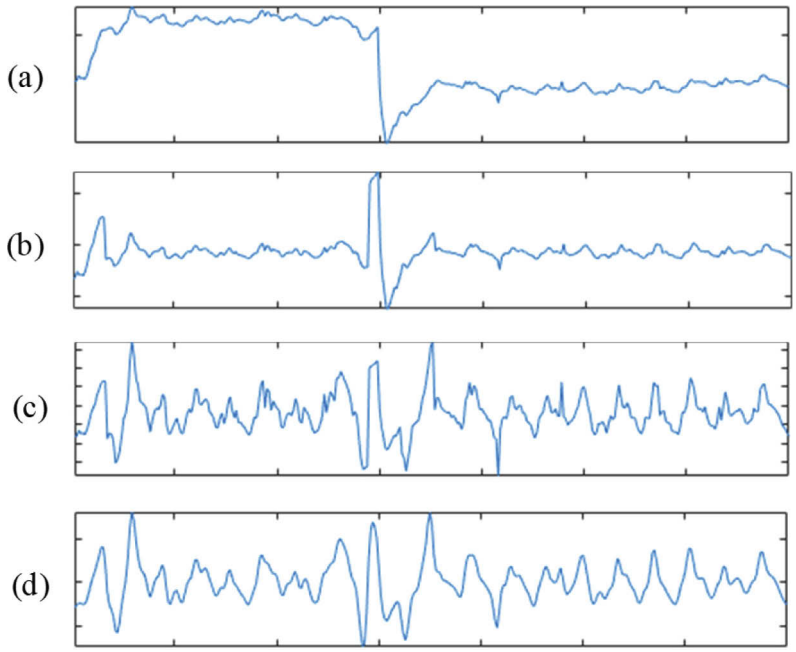

Fig. 3. Baseline cancelation and smoothing. (a): the raw pulse, (b) the pulse after baseline cancelation, (c) the pulse after spike smoothing, (d) the pulse after five-point cubic smoothing.

Notice that Eq. (5), Eq. (6), Eq. (8) and Eq. (9) are used for the four points in both ends, and Eq. (7) is used for all the other points.

$$\overline{Y}_{-2} = \frac{1}{70}(69Y_{-2} + 4Y_{-1} - 6Y_0 + 4Y_1 - Y_2) \tag{5}$$

$$\overline{Y}_{-1} = \frac{1}{30}(2Y_{-2} + 27Y_{-1} + 12Y_0 - 8Y_1 + 2Y_2) \tag{6}$$

$$\overline{Y}_0 = \frac{1}{35}(-3Y_{-2} + 12Y_{-1} + 17Y_0 + 12Y_1 - 2Y_2) \tag{7}$$

$$\overline{Y}_1 = \frac{1}{35}(2Y_{-2} - 8Y_{-1} + 12Y_0 + 27Y_1 + 2Y_2) \tag{8}$$

$$\overline{Y}_2 = \frac{1}{70}(-Y_{-2} + 4Y_{-1} - 6Y_0 + 4Y_1 + 69Y_2) \tag{9}$$

Five-point cubic smoothing algorithm utilizes polynomial least square to approximate sampling points, it can produce much better results than conventional moving average filter in detail-preserving. The signal after five-point cubic smoothing is shown as Fig. 3(d).

2.3. Self-adaptive singular spectrum analysis

A new method that combines singular spectrum analysis with self-adaptive function (self-adaptive SSA) is proposed to remove irregular noise and extract clean pulse wave with useful details reserved. Self-adaptive SSA is a global analysis method based on phase space reconstruction which decomposes original signal into multiple variable components [27] and choose appropriate components automatically to reconstruct pulse wave. The main steps of the proposed self-adaptive SSA method include trajectory matrix construction, singular value decomposition, self-adaptive components selection, and signal reconstruction. Through the use of self-adaptive SSA, main periodic components in original signal could be extracted to reconfigure pulse wave.

Biomedical Optics EXPRESS

1) Trajectory matrix construction: Suppose the total length of time series *x(t)* is *N*, the window length used to construct trajectory matrix is *L*, and define $K = N\text{-}L + 1$, then the trajectory matrix can be represented as:

$$\mathbf{X} = \begin{bmatrix} x_1 & x_2 & \cdots & x_K \\ x_2 & x_3 & \cdots & x_{K+1} \\ \vdots & \vdots & \vdots & \vdots \\ x_L & x_{L+1} & \cdots & x_N \end{bmatrix} \tag{10}$$

It is obvious that the trajectory matrix is a Hankel matrix. The main concern here is the choose of window length *L*, and according to Mahmoudvand et al. [28], when *L* takes the median value of *N*, the corresponding singular value will get the maximum, so the literature recommended that in most cases the median value of *N* is the best choice for *L*.

2) Singular value decomposition: This method has already been widely used in principal component analysis (PCA), pattern recognition, data compression, and so on. The singular value decomposition of trajectory matrix $\mathbf{X}$ is given as:

$$\mathbf{X} = \sum_{i=1}^{d} \lambda_i \mathbf{U}_i \mathbf{V}_i^{\mathrm{T}} \tag{11}$$

where d is the number of non-zero singular values of $\mathbf{X}$, and it is obvious that $d = rank(\mathbf{X}) \le min(L, K)$. Here, λ_1 is the largest singular value, and the value of λ_i decreases as the index i increases, this means components with lower index number contributes to the origin signal more. $\mathbf{U}_\mathrm{i}$ and $\mathbf{V}_\mathrm{i}$ are the left and right singular vectors of $\mathbf{X}$ respectively.

3) Self-adaptive components selection: It is generally believed that the main energy of a signal is concentrated on the first r ($r < d$) larger singular values, while the smaller singular values are regarded as noise components. Figure 4 shows an example of original signal and its 5 components reconstructed from the former 5 singular values respectively, and it is obvious that the first component has the highest energy, while the last component has the lowest energy. The purpose of components selection is to determine an appropriate value for r, so that the noise-free pulse wave can be reconstructed from the former r singular values. If r is too small, the useful details like dicrotic wave will be lost, and if r is too large, the noise reduction performance will be poor, thus the exact augmented Lagrange multiplier algorithm (EALM) is adopted to achieve self-adaptive components selection [29]. High-dimensional data can be considered sparse, and its sparsity is reflected in its low rank property, thus trajectory matrix $\mathbf{X}$ can be approximated by its best low rank matrix $\mathbf{S}$. Suppose $\mathbf{X}$ is contaminated by slight Gauss noise $\mathbf{E}$ and sparse big noise $\mathbf{W}$, and $||\mathbf{E}||_\mathrm{F} \le \delta$ where $||\mathbf{E}||_\mathrm{F}$ represents the nuclear norm of $\mathbf{E}$, then matrix denoise problem can be described by the following optimization problem:

$$\begin{aligned} &\min \parallel \mathbf{S}\parallel_* + \gamma \parallel \mathbf{W}\parallel_0, \gamma = 1/\sqrt{max(m,n)} \\ &s.t. \parallel \mathbf{X} - \mathbf{S} - \mathbf{W} \parallel \le \delta \end{aligned} \tag{12}$$

where m, n denotes the number of rows and columns of matrix $\mathbf{X}$ respectively, $||\mathbf{W}||_0$ denotes the zero-norm of matrix $\mathbf{W}$. To solve this optimization problem, the following Lagrange function is

Biomedical Optics EXPRESS

defined:

$$\begin{aligned} L(\mathbf{S},\mathbf{W},\mathbf{Y},\mu) &= \|\mathbf{S}\|_* + \gamma \parallel \mathbf{W}\|_0 + \mathbf{Y}^T(\mathbf{X}-\mathbf{S}-\mathbf{W}) + \frac{\mu}{2} \parallel \mathbf{X}-\mathbf{S}-\mathbf{W}\|_F^2 \\ &= \parallel \mathbf{S}\|_* + \gamma \parallel \mathbf{W}\|_0 + \mathbf{Y}^T H(\mathbf{X}) + \frac{\mu}{2} \parallel H(\mathbf{X})\|_F^2 \end{aligned} \tag{13}$$

where μ is the punishment coefficient, and the initial value of Lagrange multiplier $\mathbf{Y}$ can be represented as:

$$\mathbf{Y}_0 = sgn(\mathbf{X})/max(\|sgn(\mathbf{X})\|_2, \gamma^{-1} \parallel sgn(\mathbf{X})\|_\infty) \tag{14}$$

where sgn is signum function, $||\mathbf{X}||_\infty$ denotes the infinite-norm of $\mathbf{X}$, and $||\mathbf{X}||_2$ denotes the 2-norm of $\mathbf{X}$. The following EALM algorithm could be used to iteratively solve Eq. (13):

EALM Algorithm: compute the best low rank matrix $\mathbf{S}$ of original trajectory matrix $\mathbf{X}$.

Input: Observation matrix $\mathbf{X}$.

Initialization: $\mathbf{Y}_0$ shown in Eq. (14), $\mathbf{W}_0$=0, 0<μ_0<1, ρ>1, k=0;

while not converged **do**

solve
$$\mathbf{S}_{k+1} = \arg\min_{\mathbf{S}}(L(\mathbf{S},\mathbf{W}_k,\mathbf{Y}_k,\mu_k))$$
$$\mathbf{W}_{k+1} = \arg\min_{\mathbf{W}}(L(\mathbf{S}_{k+1},\mathbf{W},\mathbf{Y}_k,\mu_k))$$
$$\mathbf{Y}_{k+1} = \mathbf{Y}_k + \mu_k H(\mathbf{X})$$
$$\mu_{k+1} = \rho\mu_k$$
k=k+1

end while

RETURN: $\mathbf{S}_k$

The convergence condition of EALM algorithm is that $L(\mathbf{S}, \mathbf{W}, \mathbf{Y}, \mu)$ derives $\mathbf{S}$ equals to zero and $H(\mathbf{X})\leq\delta$. Reference [30] proved theoretically that the Lagrange multiplier $\mathbf{Y}$ is sufficient to guarantee the linear convergence of the EALM algorithm when the function $H(\mathbf{X})$ is continuously differentiable. When the algorithm converges, the output matrix $\mathbf{S}_k$ is the best low rank matrix to approximate the source matrix $\mathbf{X}$, and it is recommended to reconstruct pulse wave with the first $rank(\mathbf{S}_k)$ singular values, in other words, the appropriate value for r is $rank(\mathbf{S}_k)$. Notice that EALM algorithm requires μ to iterate from a smaller positive number so that noise matrix $\mathbf{W}$ can be well estimated. For an extreme counterexample, if the initial value μ_0 is infinite, the algorithm converges in the first iteration and the noise matrix $\mathbf{W}$ is not estimated at all. The coefficient ρ determines the convergence rate which should be set based on the balance between time and accuracy of the algorithm, and the typical value for ρ is between 1.1 and 2.

4) Signal reconstruction: Signal can be reconstructed from the former r larger singular values using the matrix **RCA** shown in Eq. (15). **RCA** is also called reconstruction matrix, which is defined as the product of first r columns of $\mathbf{U}$ and first r rows of $\mathbf{V}^{\mathrm{T}}$.

$$\mathbf{RCA} = \mathbf{U}(1,2,\cdots,r) * \mathbf{V}^{\mathrm{T}} \begin{pmatrix} 1 \\ 2 \\ \vdots \\ r \end{pmatrix} \tag{15}$$

Biomedical Optics EXPRESS

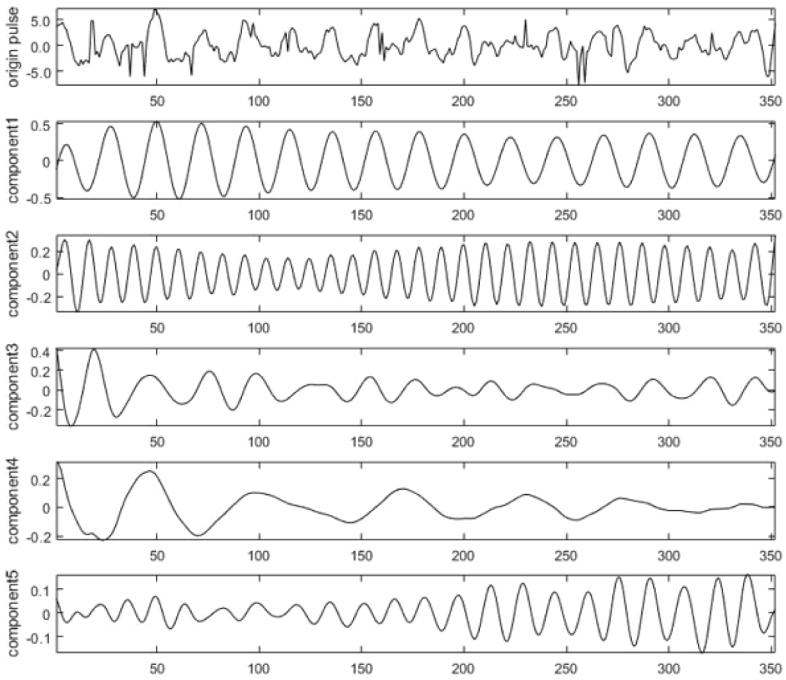

Fig. 4. Original pulse and its main components 1-5.

The dimension of matrix **RCA** is $L\times K$, and furtherly define $L^*=\min(L, K)$, $K^*=\max(L, K)$, and $y_{i,j}$ represents the value of column j in row i of matrix **RCA**, then the reconstructed signal y_{rc} can be obtained by:

$$y_{rc}=\begin{cases}\frac{1}{k}\sum\limits_{m=1}^{k} y_{m,k-m+1} & 1\le k<L^*\\ \frac{1}{L^*}\sum\limits_{m=1}^{L^*} y_{m,k-m+1} & L^*\le k\le K^*\\ \frac{1}{N-k+1}\sum\limits_{m=k-K^*+1}^{N-K^*+1} y_{m,k-m+1} & K^*<k\le N\end{cases} \tag{16}$$

After the above steps, the denoised pulse wave could finally be obtained and Fig. 5 shows the actual performance of the proposed self-adaptive SSA method. Here, the total length of pulse signal is 160, the window length L is set to 80, and the number of singular values used to reconstruct pulse wave is 8 ($r=8$). After using self-adaptive SSA, the irregular noise caused by motions or illumination changes are reduced significantly, while useful details like dicrotic wave preserved.

Biomedical Optics EXPRESS

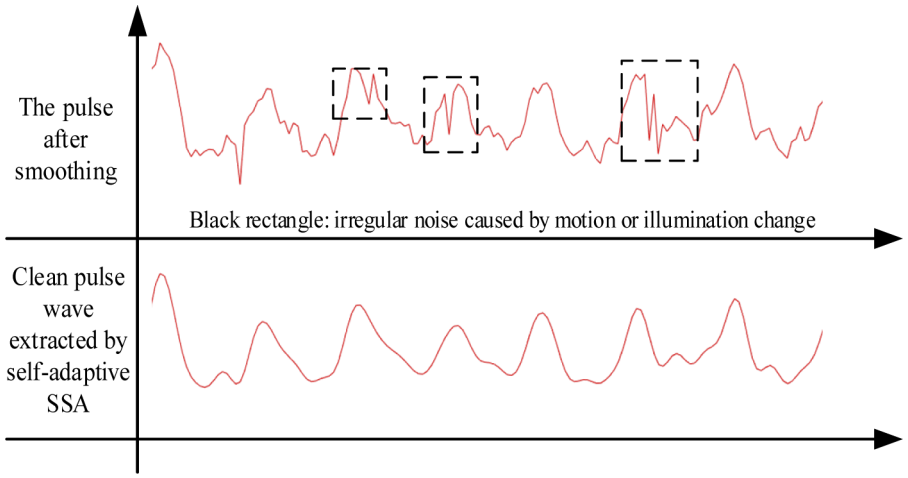

Fig. 5. Extract clean pulse wave using self-adaptive SSA.

3. Experiments

3.1. Experimental setup

MAHNOB-HCI is a public database which contains videos of 30 subjects recorded at 61 fps with a resolution of 780×580 pixels, and the synchronized ECG signals are provided as the ground truth [19]. Our method is evaluated on MAHNOB-HCI database for the following reasons: 1) it contains abundant facial videos recorded under realistic situations, and both illumination variations and subjects' motions are involved; 2) ground-truth HR values are recorded by ECG simultaneously; 3) it is a public database that can be easily accessed by all researchers who would like to do further comparison fairly [19].

The performance of our method as well as other state-of-the-art methods [9,11,16,17,18] are compared on average HR, and real time HR (RTHR) using MAHNOB-HCI database. One video corresponds to one average HR calculated by our improved peak detection algorithm which uses both amplitude and time thresholds to pick out main peaks in pulse wave and compute the mean value of peak to peak intervals for HR estimation. RTHR values are obtained from a 5-second sliding window (4-second overlap) also using our improved peak detection algorithm.

The Pearson correlation coefficient, pulse rate variability (PRV), ascending limb precision, descending limb precision, dicrotic duration precision and dicrotic amplitude precision are used as the indicators to evaluate detail preserving ability. Here, ground-truth pulse waves are recorded from earlobe at 200 Hz using a PPG sensor, while facial videos are recorded at 30 fps simultaneously using a consume-level camera of Logitech.

Furthermore, our method is also evaluated on smartphone to verify its accuracy in practice. The video recording parameters for smartphone is set to 30 fps with a resolution of 1280×720 pixels, and the ground truth HR are recorded simultaneously using the pulse oximeter of Heal Force.

3.2. Experimental results on the MAHNOB-HCI database

MAHNOB-HCI database contains abundant facial videos recorded under realistic situations, and both illumination variations and subjects' motions are involved which makes pulse wave extraction challenging. In this section, 208 videos are picked from MAHNOB-HCI database for experiments, and the time duration of each video varies from 13 seconds to 20 seconds.

Biomedical Optics EXPRESS

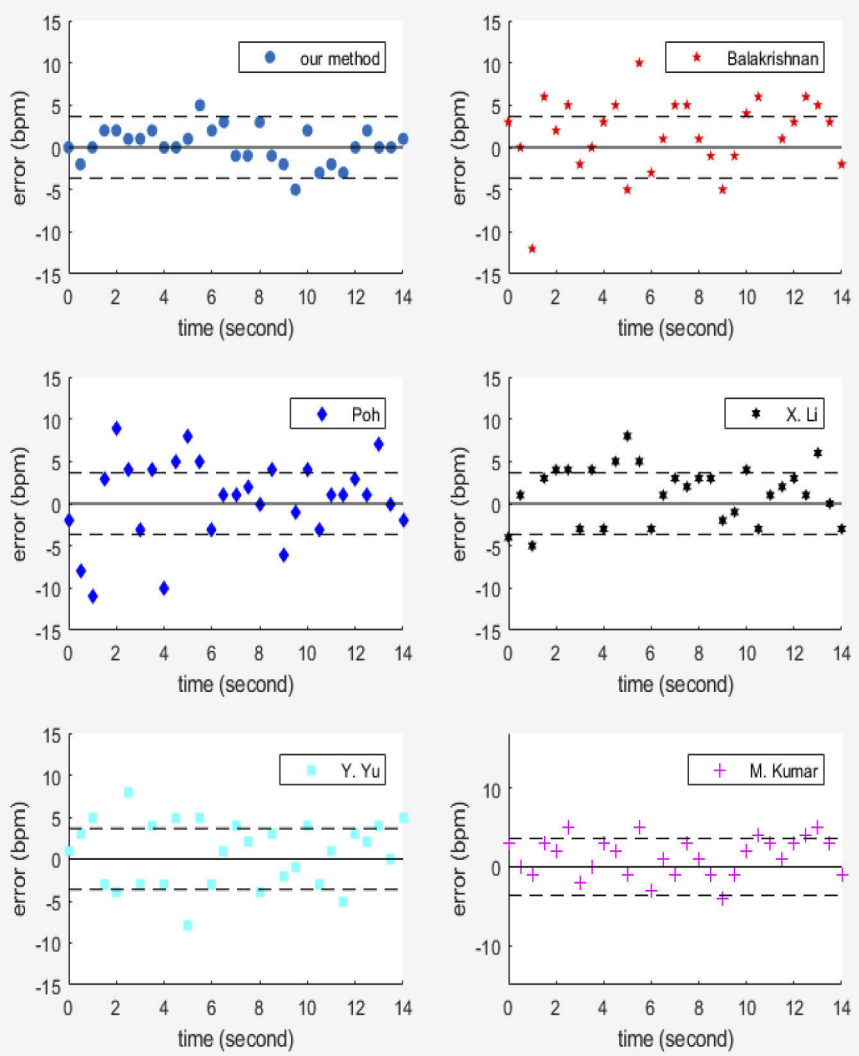

Fig. 6. Bland-Altman diagrams of our proposed method and other state-of-the-art methods. Each subplot gives the mean error between estimated RTHR sequence and ground truth RTHR sequence, and the ground truth RTHR sequence is obtained from synchronized ECG signal.

In order to evaluate the performance of RTHR estimation, a 5-second sliding window (4-second overlap) is adopted to get RTHR sequence, and the consistence between the estimated RTHR sequence and ground truth RTHR sequence is assessed using Bland-Altman plot as shown in Fig. 6 where our method and other state-of-the-art methods are compared together. The two dotted lines in each subplot represent the confidence range [μ-1.96σ, μ+1.96σ], and only the points between the dotted lines are considered highly credible. Results in Fig. 6 show that our method outperforms other state-of-the-art methods in RTHR estimation with 2 values slightly out of bounds. The result of Kumar et al. [18] is also acceptable with 5 values slightly out of bounds.

The proposed method is also compared with other state-of-the-art methods on average HR estimation, and Table 1 shows the mean absolute error (MAE), standard deviation (SD), root mean square error (RMSE) and Pearson correlation coefficient ρ between estimated HR and

Biomedical Optics EXPRESS

ground truth. Experimental results show that our method achieves superior performance to the methods in [9,11,16,17] and similar to that of Kumar et al. [18] under realistic situations in terms of average HR estimation.

Table 1. Performance comparison on average HR estimation using MAHNOB-HCI database

Method	MAE (bpm)	SD (bpm)	RMSE (bpm)	ρ
Balakrishnan [9]	8.71	24.30	25.12	0.81
Poh [11]	4.16	14.52	15.43	0.83
X. Li [16]	3.67	9.48	11.64	0.92
Y. Yu [17]	5.94	11.00	9.33	0.86
M. Kumar [18]	2.47	5.13	5.40	0.94
Our method	2.05	4.92	5.27	0.96

Furthermore, the performance comparison for different skin tones is also studied. MAHNOB-HCI database contains samples of different countries which covers nearly all kinds of skin tones and thus can be divided into white, olive and black according to the shade of color. 12 subjects are picked out (4 white, 4 olive and 4 black) for tests and Fig. 7 shows the agreement between measured and ground-truth RTHR of different skin tones. Although the white skin tone category got the best agreement, the proposed method performs well for the other two categories too, which verified its effectiveness in processing low SNR signals.

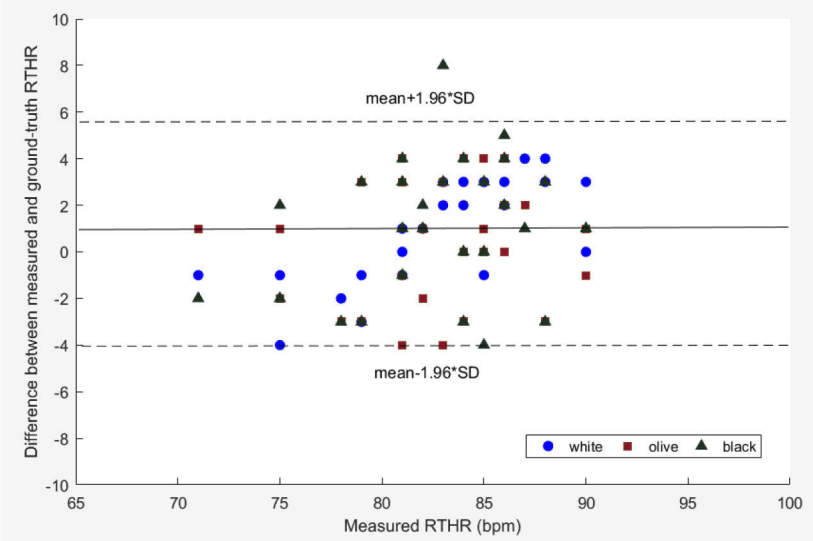

Fig. 7. Comparison of HR measured from consumer-level camera and from ground truth pulse oximeter using the proposed method for different skin tones: white, olive and black.

3.3. *Experimental results of detail-preserving capability*

As for detail-preserving ability, the extracted pulse wave of our method is compared with three state-of-the-art methods [16,17,18]. Methods in [9,11] are both based on blind source separation which combines signals in red, green and blue channels to get one independent component, but any interference would affect the three channels at the same time, thus methods in [9,11] would unlikely be able to recover pulse wave when head motions or illumination changes are involved. Therefore, we believe our method outperforms [9,11] in real mobile scenarios and

Biomedical Optics EXPRESS

thus not making experimental comparison with them. Figure 7 is the test scenario for evaluating detail-preserving ability.

In this experiment, 5 individuals were tested (3 males and 2 females) and each subject received 10 tests with slightly head motions and facial expression changes permitted. The calculated Pearson correlation coefficient is used to evaluate pulse recovery ability of each method, and the higher the coefficient value, the more accurate the extracted pulse wave is. The root mean square error (RMSE) of PRV estimation is also compared to show how accuracy the method could achieve. Ascending limb precision (P_u), descending limb precision (P_d), dicrotic duration precision (P_t) and dicrotic amplitude precision (P_h) are defined in Eq. (17), Eq. (18), Eq. (19) and Eq. (20) respectively.

$$P_u = \frac{1}{n}\sum |U_p - U_g| \tag{17}$$

$$P_d = \frac{1}{n}\sum |D_p - D_g| \tag{18}$$

$$P_t = \frac{1}{n}\sum |T_p - T_g| \tag{19}$$

$$P_h = \frac{1}{n}\sum |AR_p - AR_g| \tag{20}$$

where n is the number of tests for each subject. U_p and U_g represent the time duration of ascending limb of extracted pulse wave and ground-truth respectively, while D_p and D_g represent the time duration of descending limb of extracted pulse wave and ground-truth. T_p and T_g are the time duration between systolic peak and dicrotic peak of extracted pulse wave and ground-truth respectively. AR_p and AR_g are the amplitude ratio of systolic peak to dicrotic peak of extracted pulse wave and ground-truth. Note that, the single period waveforms with no dicrotic wave will be discarded when calculating P_t and P_h. It is evident that P_u and P_d are related to the accuracy of overall shape, while P_t and P_h can characterize the location and amplitude accuracy of dicrotic wave respectively. Table 2 shows the test result of each method and Fig. 9 gives an example of extracted facial pulse wave using each method.

Table 2. Experimental results of detail-preserving capability. Comparison with state-of-the-art methods.

Method	Pearson correlation coefficient	RMSE of PRV estimation (ms)	P_u (ms)	P_d (ms)	P_t (ms)	P_h
X. Li	0.66	64	36	47	22	0.72
Y. Yu	0.54	75	33	51	30	1.15
M. Kumar	0.83	26	27	35	14	0.23
The proposed method	0.91	23	19	28	11	0.16

Experimental results in Table 2 and Fig. 8 demonstrated that our method performs best in detail-preserving with relevance of 0.91 and smaller values for the parameters RMSE, P_u, P_d, P_t and P_h achieved. Due to the usage of narrowband filter, the pulse wave extracted by Yu et al. approaching sine wave. Method of Li et al. is acceptable with inapparent dicrotic wave reserved. The extracted pulse wave of M. Kumar has obvious dicrotic waves but exists distortion compared with ground-truth.

3.4. Experimental results on smartphone

Our algorithms were also transplanted to mobile phone (HUAWEI P20) and 5 individuals were tested (3 males and 2 females). Everyone was asked to do two sets of tests: one for stationary test

Biomedical Optics EXPRESS

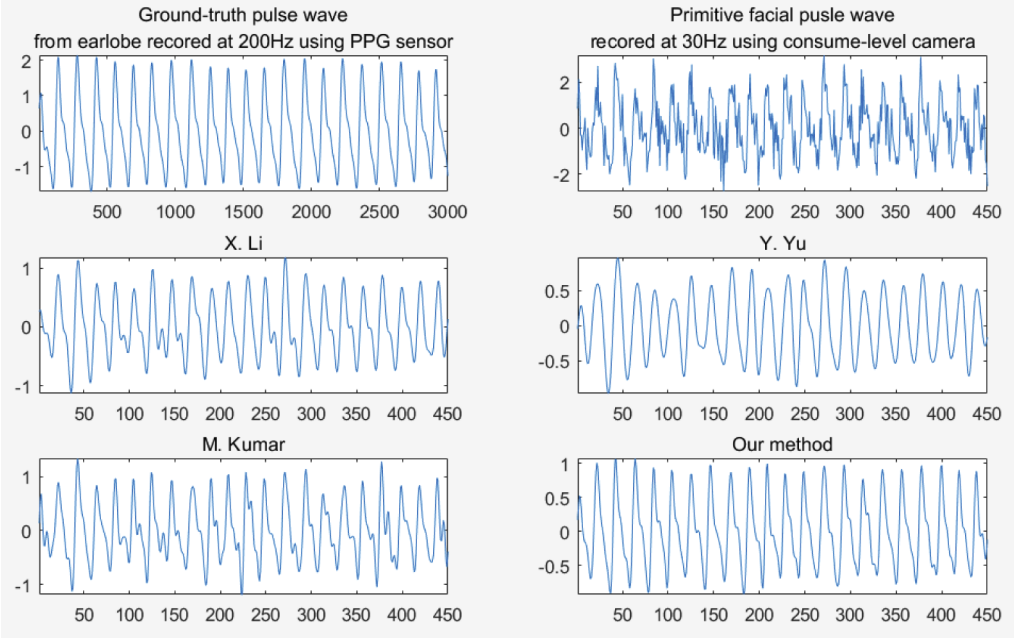

Fig. 8. Comparison of extracted pulse wave from facial video (a high SNR example).

where the face of subject keeps motionless, and another for dynamic test where the subject is free to speak and move head. Each subject received 10 stationary and dynamic tests respectively, and the ground truth values are recorded by pulse oximeter of Heal Force simultaneously. Figure 9 shows how the subjects were tested.

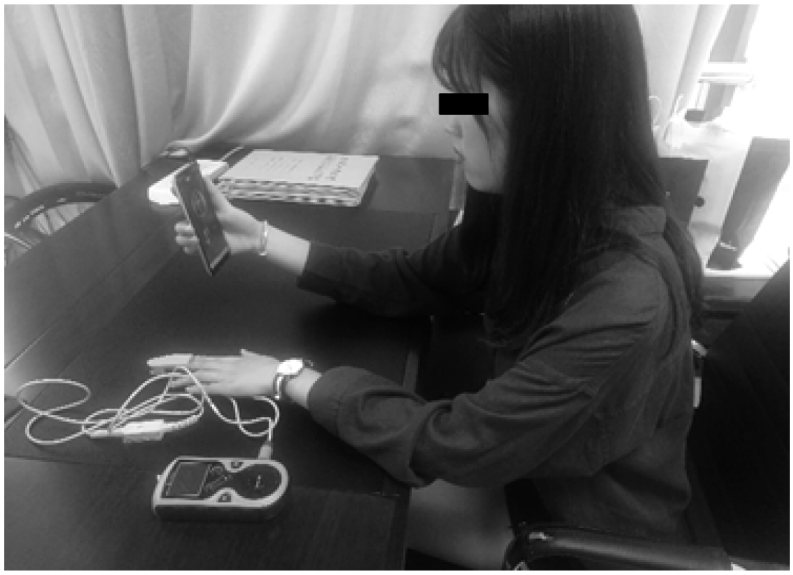

Fig. 9. Test scenario for HR estimation using smartphone.

Biomedical Optics EXPRESS

Experimental results are shown in Table 3, our method achieved satisfying accuracy of HR estimation on smartphone in both stationary and dynamic scenarios. For stationary scenario, the mean absolute error (MAE) of tested subjects is lower than 2 bpm which reaches medical standard, and for dynamic scenario, the MAE is about 3 bpm which is acceptable for daily HR monitoring. The developed android APP will be online soon.

Table 3. HR estimation based on smartphone using the proposed method

category	Subject ID	MAE (bpm)	SD (bpm)	RMSE (bpm)
stationary	1	1.03	2.33	3.12
	2	1.70	3.07	2.94
	3	0.58	1.41	1.53
	4	2.21	3.68	3.43
	5	1.12	2.06	3.70
dynamic	1	2.15	3.24	3.18
	2	3.41	5.26	5.52
	3	2.57	5.17	4.79
	4	2.12	3.19	4.30
	5	3.10	4.33	5.84

4. Conclusion

In this paper, a new method for detail-preserving pulse wave extraction from facial videos is introduced. The method consists of preprocessing procedure and a newly proposed self-adaptive SSA algorithm. Our experiments are based on a public database called MAHNOB-HCI so that all researchers who would like to do further comparison fairly could access this database easily. Experimental results show that the proposed method outperforms state-of-the-art methods in average HR, RTHR, and detail-preserving. The mean absolute error of the estimated HR using our method is 2.05 bpm while that of best-performance contrasted method is 2.47 bpm. As for detail-preserving capability, the pulse wave extracted by our method shows a very high correlation of 0.91 with ground-truth.

Another contribution of the proposed method is that it enabled the non-contact estimation of atrial fibrillation, heart rate variability, blood pressure, as well as other physiological parameters that require standard pulse wave on portable devices. Our approach has already been implemented on smartphone (android version) and achieved satisfactory HR estimation. The developed android APP will be online soon.

Although the proposed method meets the requirement of accurate pulse wave extraction under realistic situations, the presence of large motions, such as intense head swing, is still a remaining challenge and a large motion disturbance detection and disposal algorithm should be included in further work.

Acknowledgements

This work is supported by 2018 Training Programme Foundation for Application of Scientific and Technological Achievements of Hefei University of Technology, and is also supported by 2019 Independent Innovation Project of Industrial Safety and Emergency Technology of Anhui Key Laboratory. We sincerely thank Professor Yang and other colleagues for providing advices for solving technological problems during project research.

Biomedical Optics EXPRESS

Disclosures

The authors declare that there are no conflicts of interest related to this article.

References

1. T. Wu, V. Blazek, and H. J. Schmitt, "Photoplethysmography imaging: A new noninvasive and non-contact method for mapping of the dermal perfusion changes," Proc. SPIE **4163**, 62–70 (2000).
2. J. Allen, "Photoplethysmography and its application in clinical physiological measurement," Physiol. Meas. **28**(3), R1–R39 (2007).
3. M. Z. Poh, D. J. Mcduff, and R. W. Picard, "Non-contact, automated cardiac pulse measurements using video imaging and blind source separation," Opt. Express **18**(10), 10762–10774 (2010).
4. I. Pavlidis, J. Dowdall, N. Sun, C. Puri, J. Fei, and M. Garbey, "Interacting with human physiology," Computer Vis. Image Understand **108**(1-2), 150–170 (2007).
5. W. Liu, X. Fang, Q. Chen, Y. Li, and T. Li, "Reliability analysis of an integrated device of ECG, PPG and pressure pulse wave for cardiovascular disease," Microelectron. Reliab. **87**, 183–187 (2018).
6. W. Zhong, K. J. Cruickshanks, C. R. Schubert, C. M. Carlsson, R. J. Chappell, B. E. Klein, R. Klein, and C. W. Acher, "Pulse wave velocity and cognitive function in older adults," Alzheimer Dis. Assoc. Disord. **28**(1), 44–49 (2014).
7. H. S. James, R. Stern, H. B. Jodi, E. K. Claudia, W. H. Erika, Y. Terry, and E. P. Paul, "Relationships between sleep apnea, cardiovascular disease risk factors, and aortic pulse wave velocity over 18 years: the Wisconsin Sleep Cohort," Sleep & Breathing **20**(2), 813–817 (2016).
8. H. A. Lane, J. C. Smith, and J. S. Davies, "Noninvasive assessment of preclinical atherosclerosis," Vasc. Health Risk Manage. **2**(1), 19–30 (2006).
9. G. Balakrishnan, F. Durand, and J. Guttag, "Detecting pulse from head motions in video," *2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 3430–3437 (2013).
10. S. A. Siddiqui, Y. Zhang, Z. Feng, and A. Kos, "A pulse rate estimation algorithm using PPG and smartphone camera," J. Med. Syst. **40**(5), 126 (2016).
11. M. Z. Poh, D. J. McDuff, and R. W. Picard, "Advancements in noncontact, multiparameter physiological measurements using a webcam," IEEE Trans. Biomed. Eng. **58**(1), 7–11 (2011).
12. H. Wu, M. Rubinstein, E. Shih, J. Guttag, F. Durand, and W. Freeman, "Eulerian video magnification for revealing subtle changes in the world," ACM Trans. Graph. 31(4), 1–8 (2012).
13. R. Amelard, D. A. Clausi, and A. Wong, "A spectral-spatial fusion model for robust blood pulse waveform extraction in photoplethysmographic imaging," Biomed. Opt. Express **7**(12), 4874–4885 (2016).
14. D. Laure and I. Paramonov, "Improved Algorithm for Heart Rate Measurement Using Mobile Phone Camera," *2013 13th Conference of Open Innovations Association (FRUCT)*, 2343–0737(2013).
15. B. P. Yan, C. K. Chan, C. K. Li, O. T. To, W. H. Lai, G. Tse, Y. C. Poh, and M. Z. Poh, "Resting and postexercise heart rate detection from fingertip and facial photoplethysmography using a smartphone camera: a validation study," JMIR Mhealth Uhealth **5**(3), e33 (2017).
16. X. Li, J. Chen, G. Zhao, and M. Pietikainen, "Remote heart rate measurement from face videos under realistic situations," *2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 4264–4271(2014).
17. Y. Yu, P. Raveendran, and C. Lim, "Heart rate estimation from facial images using filter bank," *2014 6th International Symposium on Communications, Control and Signal Processing (ISCCSP)*, 69–72(2014).
18. M. Kumar, A. Veeraraghavan, and A. Sabharwal, "DistancePPG: Robust non-contact vital signs monitoring using a camera," Biomed. Opt. Express **6**(5), 1565–1588 (2015).
19. M. Soleymani, J. Lichtenauer, T. Pun, and M. Pantic, "A multimodal database for affect recognition and implicit tagging," IEEE Trans. Affective Comput. **3**(1), 42–55 (2012).
20. P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* 1, 511–518 (2001).
21. P. Sahindrakar, G. De Haan, and I. Kirenko, "Improving motion robustness of contact-less monitoring of heart rate using video analysis," *Eindhoven*, Technische Universiteit Eindhoven, Department of Mathematics and Computer Science, The Netherlands (2011).
22. W. Verkruysse, L. O. Svaasand, and J. S. Nelson, "Remote plethysmographic imaging using ambient light," Opt. Express **16**(26), 21434–21445 (2008).
23. R. Stricker, S. Muller, and H. M. Gross, "Non-contact video-based pulse rate measurement on a mobile service robot," *The 23rd IEEE International Symposium on Robot and Human Interactive Communication*, 1056–1062(2014).
24. D. Y. Chen, J. J. Wang, K. Y. Lin, H. H. Chang, H. K. Wu, and Y. S. Chen, "Image sensor-based heart rate evaluation from face reflectance using Hilbert–Huang transform," IEEE Sens. J. **15**(1), 618–627 (2015).
25. W. Chen, P. Thierry, and C. Guillaume, "A comparative survey of methods for remote heart rate detection from frontal face videos," Front. Bioeng. Biotechnol. **6**(33), 33 (2018).
26. C. Huang, X. Yang, and K. T. T. Cheng, "Accurate and efficient pulse measurement from facial videos on smartphones," *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*, (2016).
27. B. Elsner and James, "Analysis of time series structure: SSA and related techniques," J. Am. Stat. Assoc. **97**(460), 1207–1208 (2002).

Corrected: 03 Apr 2020

Biomedical Optics EXPRESS

28. R. Mahmoudvand and M. Zokaei, "On the singular values of the hankel matrix with application in singular spectrum analysis," Chilean Journal of Statistics **3**(1), 43–56 (2012).
29. Z. Lin, M. Chen, L. Wu, and Y. Ma, "The augmented Lagrange multiplier method for exact recovery of corrupted low-rank matrices," Eprint Arxiv, (2010).
30. D. Bertsekas, *Constrained Optimization and Lagrange Multiplier Method* (Academic Press, 1982).

Corrected: 03 Apr 2020